Effect Of Different Population Structures On Bias And Precision Of Genotype-By-Environment (GxE) Interaction Estimates: A Simulation Study

P. Sae-Lim[*]*, H. Komen*[*]*, A. Kause*[*]

Introduction

Many aquaculture breeding programmes distribute animal material across diverse production environments, sometimes even at global scale. Selection within a nucleus broodstock may lead to lower-than-expected genetic gains in other production environments when genotype-by-environment (GxE) interaction exists but it is not introduced to the selection criteria. GxE interaction is defined as a phenomenon that genotypes response differently to an environment gradient. The most severe form of GxE interaction is re-ranking which means that ranking of genotypes changes across different environments (Lynch and Walsh (1998)). Re-ranking across environments can be estimated using a genetic correlation (r_g) between a trait measured in two environments (Falconer (1952)). GxE interaction is commonly considered to be biologically significant when genetic correlation is lower than 0.8 (Roberson (1959)). When r_g is less then 0.8, it would be useful to include both environments into a breeding programme. To accurately estimate r_g between environments, an optimal design needs to be established; an experimental design which produces an unbiased and precise result while using minimum testing capacity. Enlarging population size typically increases the power of a design but simultaneously increases costs. In contrast, too small population size or suboptimal population structure (number of families, family size, mating design) may potentially result in biased and inaccurate estimates. Furthermore, family differences in survival or in contribution to the whole population size result in unequal family sizes. Such unbalanced design may influence the bias and precision of r_g estimates. To our knowledge, no study has been done to assess bias and precision of estimates of GxE interaction. The present study describes the use of a stochastic simulation to construct an optimal population structure promoting unbiased and precise estimation of r_g between environments.

Material and methods

Population construction. The simulated population structure was a split-family design with two environments, where the offspring generation had trait records and their parents only contributed to the pedigree. In each environment, phenotype of an individual was calculated as $y = 0.5g_s + 0.5g_d + m + e$, where g_s and g_d are genetic values of sire and dam, respectively, each randomly sampled from a multi-trait normal distribution for genetic variance $N(0,V_G)$. m is Mendelian sampling term sampled from a normal distribution of a half of genetic variance $N(0,0.5V_G)$. e is an environmental effect sampled from a normal distribution of environmental variance $N(0,V_E)$. Phenotypic variance (V_P) was set to 1. Genetic variance

[*]*Animal Breeding and Genomics Centre, Wageningen University, The Netherlands.*

(V_G) was $V_P h^2$, and environmental variance (V_E) was $V_P(1-h^2)$. The simulated genetic covariance between a trait measured in two environments determined the degree of family re-ranking and was simulated as the mean of parental genetic effects plus half the genetic covariance (Mendelian sampling). No environmental covariance was simulated between the two environments, i.e., each animal inhabited only one environment.

Simulated scenarios. The simulated population had 1 male : 2 female mating ratio, r_g of 0.8 (threshold value), and a trait with heritability (h^2) of 0.3 or 0.1 in each environment. Three population structure scenarios were simulated: **A** = Increasing population size with 100 full sib families and full-sib family size increasing from 3 to 75 individuals in each environment. **B** = Fixed population size of 1000 in each environment but with alternative combinations of full-sib family sizes and number of families. Within each subscenario, A and B had equal family size for all families. **C** = as **B** but with unequal family sizes due to selective mortality. To create between-family variation, alternative values for survival heritability (h^2_s) and common environmental effect shared by full sibs (c^2_s) were used. The simulation was repeated 500 times for each alternative population structure.

Estimation of (co)variance components. (Co)variance components were estimated using a bivariate animal model: $y_{ij} = \mu_i + a_{ij} + e_{ij}$, where i represents a trait measured in two environments, i =1, 2 (two traits); μ_i is the overall mean of the trait i; a_{ij} is the random genetic effect individual j; e_{ij} is the random residual effect. Genetic correlation and standard error (SE) were estimated using Restricted maximum likelihood (REML) in ASReml.

Results and discussion

Downward biased estimates. A design with small family size resulted in downward biased r_g estimates. In scenario **A**, for h^2=0.1, biased correlation was found when family size was 3-20. For h^2=0.3, only a slight downward bias was obtained when family size was 3-10 (Fig.1A). In scenario **B**, biased estimates were obtained when either family size was low (with high family number) or family number was too low (with high family size) (Fig.1B). When the family size and heritability are reduced, it is increasingly difficult to estimate true family mean and thus also breeding values of individuals. This artificially increases re-ranking and lowers the estimate for r_g. Downward biased GxE correlation estimate from poor design thus indicates the existence of GxE interaction when it is not true. This can lead to an erroneous conclusion that a multi-environment breeding programme is needed . When r_g was 0.8 and h^2 was 0.1, the mean estimated r_g at family size of 25 never reached the true correlation of 0.8 (Fig.1B), but median and mode were 0.8 (data not shown). This is because distribution of estimated r_g is pushed against the limit of unity, and the distribution becomes skewed. In addition to population structure, the method of estimating r_g influences the degree of bias. For example, family-means method produces downward bias when the true genetic correlation is high (Windig (1997); Astles et al. (2006)). REML is known to be robust against small and unbalanced designs (Patterson and Thompson (1971); Harville (1977)). However, at extreme population structures, bias can occur because parameter estimation is difficult, as shown here.

Optimal design. When population size was fixed, both family size and family number were varied. Consequently the optimal design tended to be the balance of the two to reach the lowest SE and the minimum bias. Figure 1B shows that for h^2 = 0.1, the optimal design was

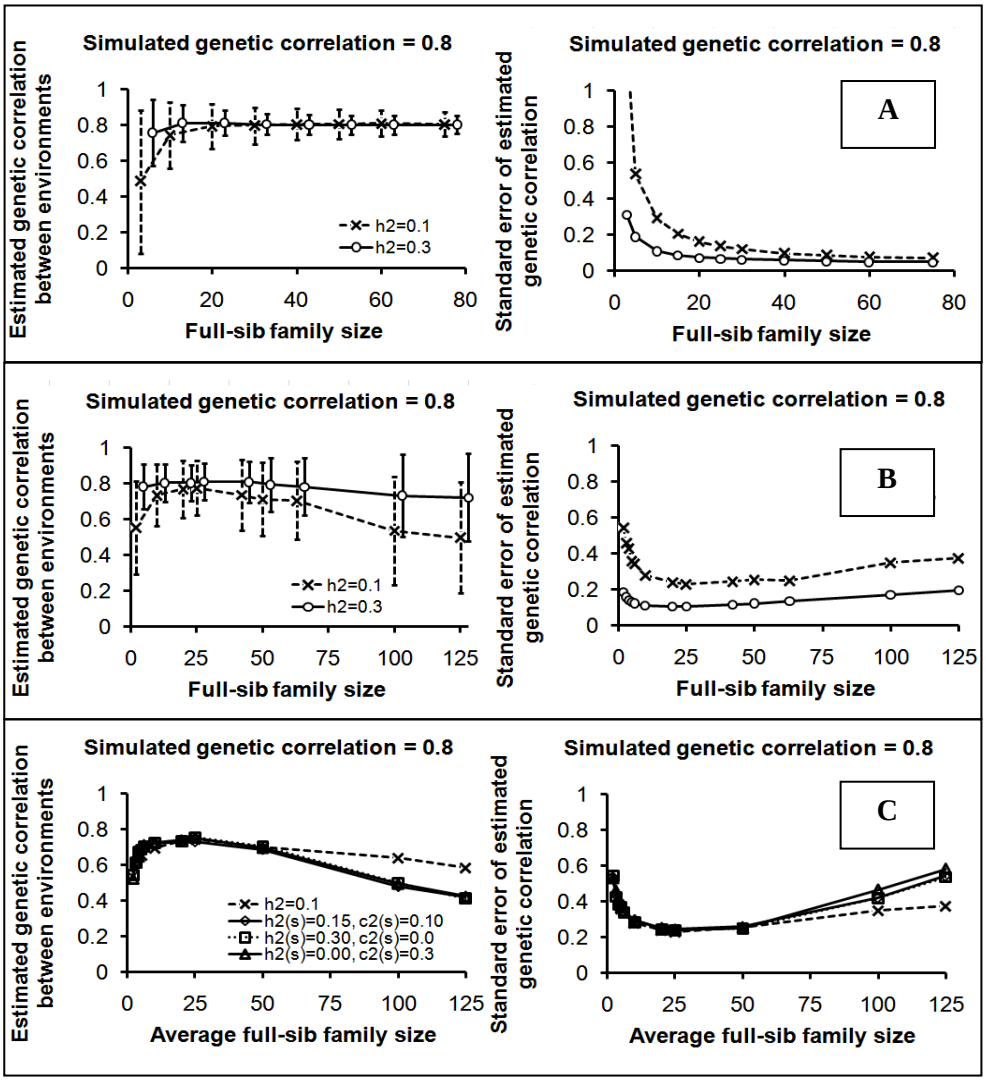

Figure 1: The simulated genetic correlation between two environments and standard error of the estimates. A= 100 families, family size 3-75; B = 1000 animals, family size 2-125 with family numbers of 500-8; C= as B but with unequal family sizes: family size is average of the families. In the estimated genetic correlation graphs, symbols for h^2 of 0.3 have been moved 3 points forward on the x-axis to distinguish the lines of h^2 of 0.1 and 0.3. Y-bar error is the standard deviation of estimated genetic correlations from all simulated replicates.

obtained when the family size was about 20-25 individuals (family number of 50-40). For h^2 = 0.3, the optimal design was obtained when the family size was about 10-25 (family number of 100-40) (similar results were obtained from mean square error; data not shown). Similar to

our results, for a rainbow trout breeding programme, an increase in accuracy of estimated breeding value of sea body weight (h^2=0.22) for a breeding candidate located at a freshwater nucleus levels out when the number of individuals tested at the sea increases from 7 to 20 recorded per family (Martinez et al. (2006)). However, when studying multiple traits, optimizing design simultaneously for low h^2 and high h^2 traits is challenging. The exception to this is the optimisation of designs for survival (low h^2 trait) and harvest growth traits (h^2 around 0.3). When initially having large family size optimised for survival, family size at harvest is smaller for harvest growth traits, but still the design may remain close to optimal for both traits. Moreover, expected low and high h^2 traits can be recorded from different number of animals using predefined animal list at trait recording.

Unequal family sizes. Estimated SE of genetic correlation from data with unequal and equal family size was the same at the optimal design (Fig. 1C). This is in contrast to the papers by Hammersley (1949) and Tallis (1959). The likely explanation is that REML approach used in our simulation is robust against unbalanced designs (Patterson and Thomson (1971); Lynch and Walsh (1998)). Thus, SE was not strongly influenced by unequal family size at the optimal design. Even when a breeding programme is started with equal family sizes, e.g. number of eggs or number of fish stocked, there is typically family-specific mortality during all life stages (Kanis et al. (1976); Vehviläinen et al. (2008)). Besides, for mass spawning species, unequal contribution of parents to the offspring population can occur (Blonk et al. (2009)). Both mechanisms lead to unequal family sizes. An optimal design for r_g is not sensitive to family differences in survival (or parental contribution). Divergence was found between equal and unequal in estimated r_g and its SE when average full-sib family size was greater than 50 with family number less than 20 for h^2 of 0.1 (Fig. 1C). The reason might be a difficulty of estimating genetic (co) variance components when the family numbers are getting smaller.

Conclusion

Our simulation study specifically focused on the within population GxE interaction where the r_g between environments is a measurement of GxE interaction. There are three factors leading to a downward biased estimate; low h^2, small family size, and low family number. An optimal design is not affected by unequal family size due to selective mortality.

References

Astles, P.A., Moore, A.J., Preziosi, R.F. (2006). *J. Evol. Biol*., 9:114-122.

Blonk, R.J.W., Komen, J., Kamstra, A. *et al*. (2009). *Aquaculture*, 289:26-31.

Falconer, D.S. (1952). *Am. Nat.*, 86:293-298.

Hammersley, J.M. (1949). *Metron*, 15:189-205.

Harville, D.A. (1977). *J. Am. Stat Assoc.*, 72:320-338.

Kanis, E., Refstie, T., Gjedrem, T. (1976). *Aquaculture*, 8:259-268.

Lynch, M., Walsh, B. (1998). Sinauer Associates, Inc., U.S.A., page 644.

Martinez, V., Kause, A., Mäntysaari, E. *et al.* (2006). *Aquaculture*, 254:182-194.

Patterson, H.D., Thompson, R. (1971). *Biometrika*, 58:545-554.

Robertson, A. (1959). *Biometrics*, 15:469-485.

Tallis, G.M. (1959). *Aust. J. Stat.*, 1:35-43.

Vehviläinen, H., Kause, A., Quinton, C. *et al*. (2008). *Genetics*, 180:507-516.

Windig, J.J. (1997). *J. Evol. Biol.*, 10:853-874.